

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

2001

Uncertain responses by five-year-olds in a memory monitoring task

Katarina Guttmanova
The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Guttmanova, Katarina, "Uncertain responses by five-year-olds in a memory monitoring task" (2001).
Graduate Student Theses, Dissertations, & Professional Papers. 5711.
<https://scholarworks.umt.edu/etd/5711>

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.



**Maureen and Mike
MANSFIELD LIBRARY**

The University of
Montana

Permission is granted by the author to reproduce this material in its entirety,
provided that this material is used for scholarly purposes and is properly cited in
published works and reports.

****Please check "Yes" or "No" and provide signature****

Yes, I grant permission

☒

No, I do not grant permission

☐

Author's Signature: Katarina Gutmann

Date: 05/24/01

Any copying for commercial purposes or financial gain may be undertaken only with
the author's explicit consent.

UNCERTAIN RESPONSES BY FIVE-YEAR-OLDS IN A MEMORY MONITORING TASK

by

Katarina Guttmannova

B.A. Slippery Rock University of Pennsylvania, 1997

presented in partial fulfillment of the requirements

for the degree of

Master of Arts

The University of Montana

February 2001

Approved by:

Wendy E. Shields PhD

Chairperson

D. J. Stull

Dean, Graduate School

5-24-01

Date

UMI Number: EP41178

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI EP41178

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Uncertain Responses by Five-Year-Olds in a Memory Monitoring Task

Director: Wendy E. Shields, Ph.D. *WES*

This study investigated memory monitoring in preschool children via the use of inferential observation of task performance. Forty five-year olds performed in a pictorial serial-probe recognition task that featured an uncertain response. Their behavioral responses without the use of self-report were examined because the literature suggests that the use of verbal reports to assess metacognition in young children has many potential limitations. Participants were given an option to escape trials of their choosing, perhaps those that yielded the most indeterminate memory traces. It was hypothesized that five-year-olds would accurately monitor their short-term memory and the pattern of their escape responses would mirror the serial position curve of their primary memory performance, as has been documented in adult humans and rhesus monkeys. However, the pattern of responding in our five-year-olds differed from that of the adult humans and non-human primates. Children's primary memory performance was reflected in a pronounced serial position curve that corresponds to the curves found in adult human and monkey studies but the pattern of their escape responses was very different. It appears that the five-year-olds use disparate strategies when facing uncertainty. These strategies, as well as implications for future research in this area are discussed.

Table of Contents

<u>Title</u>	<u>Page Number</u>
Introduction	
Perils with the Definition	2
Previous Research on Metacognition in Children	4
Comprehension monitoring	4
Knowledge about memory	6
Memory monitoring	8
Methodological Issues	10
The Uncertainty Paradigm	14
Support for Using the SPR Task	16
Method	
Participants	18
Apparatus	19
Procedure	19
Stimulus presentation phase	19
Response phase	20
Pilot Study	22
Results	
Primary Memory Performance	23
Memory Monitoring and the Use of the Uncertain Response	24
Signal Detection Analysis of Performance	24
Sensitivity	25
Response bias	26
Discussion	27
References	34
Appendix	39
Tables	41

Figure Captions	44
Figures	45

List of Tables

Table 1. Percentage of Correct Responses for List Length 10	41
Table 2. Percentage of Correct Responses for the Split-Half Data ...	42
Table 3. Signal-Detection Distribution of Responses	43

List of Figures

Figure 1A. Serial probe recognition performance for the experimental group	45
Figure 1B. Serial probe recognition performance for the control group	46
Figure 2A. Serial probe recognition performance on the 1st half of the trials for the experimental group	47
Figure 2B. Serial probe recognition performance on the 2nd half of the trials for the experimental group	48
Figure 2C. Serial probe recognition performance on the 1st half of the trials for the control group	49
Figure 2D. Serial probe recognition performance on the 2nd half of the trials for the control group	50
Figure 3. Change in mean sensitivity for the first and second halves of sessions	51

Uncertain Responses by Five-Year-Olds in a
Memory Monitoring Task

Cognitive psychological literature was first introduced to the term "meta" in relation to memory in 1971, when John Flavell coined the term "metamemory" as one's knowledge of one's own memory (Flavell, 1971). The study of metamemory or metacognition (cognition about cognition) has been for decades an exciting albeit problematic area of research. In many of his works, Flavell (e.g., 1979, 1981, and 1987) attempted to construct an appropriate taxonomy of the concept of metacognition. In particular, he made a distinction between two aspects of metacognition: metacognitive knowledge and metacognitive experiences.

Metacognitive knowledge refers to the portion of knowledge one has available in long-term memory regarding facts about cognition in general and about one's own cognition in particular. People can often verbalize this information. Hence metacognitive knowledge is considered declarative. For example, John Smith knows that it is impossible for a person to perfectly recall an 18-digit number in a serial recall task. In fact, he knows that, in his own case, recalling an 8-digit number would be an unprecedented feat.

Metacognitive knowledge includes three different clusters of variables: person, task, and strategy. Person variables refer to the knowledge and beliefs humans have about themselves as cognitive beings (e.g., John Smith knows the capacity of his own short-term memory). Task variables correspond to the knowledge of goals or objectives of a cognitive activity (e.g., John knows that the task requires the recall of an 18-digit number). Finally, strategy variables refer to cognitions or other behaviors utilized to achieve these goals (e.g., John knows that he should use the "Palm Pilot" his mother gave him for his birthday).

Metacognitive experiences are any conscious cognitive or emotional experiences that an individual encounters before, during, or after a cognitive activity. There are two aspects of metacognitive experiences: monitoring and control (or regulation). Cognitive monitoring is the ability to assess one's own cognitive processes. It includes memory, comprehension and reality monitoring. Memory monitoring (e.g., Hart, 1965) includes assessments of one's memory with regard to the presence or absence of certain information. Comprehension monitoring (e.g., Markman, 1977) involves the realization that one does or does not understand incoming information. Reality monitoring (e.g., Johnson & Raye, 1981) refers to the ability to distinguish between dreaming, imagining and reality.

Cognitive control or regulation is the action taken to affect the course of cognitive processing. This action is the result of information obtained from the monitoring process. For example, if the monitoring of comprehension during reading indicates a lack of understanding, one might decide to switch from "automatic reading pilot" to a more effortful, attentive style of reading.

Perils with the Definition

From its inception, metacognition has been a difficult and rather complex area of study. On several occasions metacognition has been referred to as a fuzzy and inadequately understood concept (Brown, 1987; Wellman, 1983). I will now discuss some of the issues that have troubled the scientific community with regard to this topic.

First, the fact that metacognition has been defined as both knowledge about cognition and the monitoring/regulation of cognition creates a problem for the interpretation of research findings. These two components of metacognition are interconnected. However, their nature and course of development might be quite different. Knowledge about cognition, declarative knowledge, is presumed to be definable and

stable, and is usually documented rather late in development (Brown, 1987). The monitoring and regulation of cognition, on the other hand, may make use of procedural knowledge, which often is not statable. Brown suggested that monitoring and regulation of cognition might be dependent primarily on the task and circumstances and therefore could be relatively age-independent. Some researchers have argued that labeling both a person's knowledge about cognition and his or her use of that knowledge with one term is ill-advised. Hence the two-component definition should be reduced, leaving out the monitoring and regulatory processes (Cavanaugh & Perlmutter, 1982). To date, there is no general agreement on this definitional issue (Baker, 1994).

The second controversy in the literature deals with the appropriate distinction of what can be considered "meta" and what should be labeled as "cognitive" (Brown, 1987). The literature is confusing (Baker, 1994). Flavell (e.g., 1979) distinguished between these two, stating that cognitive progression is achieved via the use of cognitive strategies, whereas the monitoring of this progress employs metacognitive strategies. For example, the use of mnemonics to memorize the names of students in my class would be a cognitive strategy, and the use of mnemonics to monitor or examine my memorization enterprise (as in a self-test of memory) would be a metacognitive strategy.

The third disagreement is the argument whether metacognition can be both conscious and unconscious (Baker, 1994). This brings us back to the original Flavellian taxonomy of metacognition, namely that this concept consists of metacognitive knowledge and metacognitive experiences. Metacognitive experiences can refer to the cognitions or feelings that are easily accessed by an individual, such as the anticipation of a problem before reading a chapter on quantum physics. However, this experience can also refer to feelings that may be on the

brink of consciousness, such as a fleeting feeling of delight after accomplishing a cognitive task (Gombert, 1993).

Moreover, Brown (1987) noted that metacognitive knowledge is conscious and therefore should be declarative or definable, but an individual is assumed to have this knowledge only if he or she can state or define it. Therefore, this criterion is circular. In addition, Brown distinguishes between self-regulation during learning, which is often an unconscious experience, and conscious manipulation of one's thoughts, which is thinking about thinking or conscious regulation. She states that this distinction has been confused in the metacognition literature. The former refers to the active procedure of trial-and-error (systematic error detection and correction) evident even in very young children's learning. However, unless these behaviors reflect deliberate strategic action, they should not be termed "metacognitive". The latter corresponds to Piagetian reflected abstraction that emerges at the stage of formal operations, and allows one to solve problems or test hypotheses in thought or on a mental plane, which is obviously a very conscious experience. Therefore, Brown concluded: "Whatever distinctions must be made in order to render metacognition a more malleable concept, this one is a fine candidate for inclusion in the list" (p. 96).

Previous Research on Metacognition in Children

The first attempts to assess metacognition in children were in the area of memory (metamemory--Flavell, Friedrichs & Hoyt, 1970). Since then, metacognition has been studied in many other areas of cognition, for example, comprehension (Markman, 1977; Revelle, Wellman & Karabenick, 1985), communication (Shatz, 1978), reasoning (Kuhn, 1989), and reading (Garner, 1987).

Comprehension monitoring. Markman (1977) conducted a study on comprehension monitoring in first, second, and third graders. In her experiment, she used two tasks--a card game and a magic trick. The

children were presented with insufficient information about these tasks. For children in one group the tasks were described and demonstrated, for children in the other group the tasks were only described. The results of this study showed that children were more likely to question the instructions when the task was demonstrated, and the older children realized the insufficiency of the information before the younger children did. The first graders had almost no awareness of the fact that their comprehension of the tasks was flawed. Markman (1977) suggested that the youngest children in this study processed the information on a very superficial level, without mentally performing the instructions, and they became aware that the information was inadequate only after they were prompted to repeat the instruction or even actually perform the tasks. She suggested that the children processed the instructions more carefully, and were able to detect their comprehension failure only after they made an effort to execute the task.

A flaw in such reasoning was indicated by Revelle et al. (1985). They suggested that the younger children's requests for further explanation, which were stated only after the unsuccessful attempts to enact the instructions, did not necessarily reflect comprehension monitoring, but could have just been a reaction to their inability to execute the task. Requests for additional information could be considered strong evidence for comprehension monitoring only if they occurred before the actual attempts to perform the task. On the other hand, the absence of verbal requests for clarification does not necessarily indicate the absence of comprehension monitoring. Perhaps Markman's (1977) study would have benefited from the employment of additional measures which would evaluate comprehension monitoring, such as assessment of nonverbal behavior.

Revelle et al. (1985) conducted a study that employed a quasi-naturalistic method to examine comprehension monitoring in children 31-

51 months of age. They combined experimental methods with naturalistic settings. Experimenters' requests were standardized and structured allowing for accurate analysis of children's responses, while task and circumstances were simple and children were accustomed to them. Moreover, the methodology was further improved by utilization of a performance task as opposed to a judgment task. Performance tasks evaluate children's detection of comprehension difficulty after being exposed to questionable messages. Judgment tasks, on the other hand, require more explicit judgments of the quality of a message, such as by specifying whether the message was "faulty" or not. This latter task was used, for example, in Markman's (1977) study. The experiment of Revelle et al. involved two play-sessions--one was playing in a sandbox and the other was a pretend tea party. During these play-sessions children were presented with a variety of requests, some of which posed compliance or comprehension problems. Performance in response to these requests was recorded. The results of this study indicated that 3- and 4-year-olds possess basic comprehension monitoring skills. Participants of four years of age exhibited successful and competent monitoring of all types of problems. Discrimination between problematic and non-problematic requests, as well as proper utilization of strategies for resolving the problems were also evidenced in 3-year-olds.

This positive evidence of comprehension monitoring in very young children, which is in contrast with the general findings of earlier studies (e.g., Markman, 1977; Robinson & Robinson, 1977) illustrates the importance of choosing appropriate tools and settings when assessing metacognition in children.

Knowledge about memory. Several studies assessed children's knowledge of facts about memory by investigating their understanding of person, task, and strategy variables. The inaugural work in the area of children's metamemory was a study conducted by Flavell et al. (1970).

Although this experiment investigated children's memory monitoring, it also examined children's knowledge of person variables in respect to memory, namely the knowledge children have about their memory span. Children were asked to predict how many items they would be able to serially recall from a presented set of pictorial stimuli. Series of pictures were presented until the children labeled the sequence as too long for them to recall, or until the whole sequence was presented (10 pictures). After the prediction, the experimenter read aloud a list of familiar object names and the children were asked to recall the items serially. The difference between their actual and predicted memory span was assessed. The results suggested that older children were more accurate at predicting their memory span than younger children. The majority (64%) of 4 1/2-year old children had an unrealistic concept of their memory span. In fact, these children claimed that they would serially recall all of the items, even on the longest lists.

Wellman (1977a) researched children's understanding of additional variables that affect memory. He showed that even 3-year-olds knew that more items are harder to remember than fewer items (task variable), and some also knew that noise intervened with remembering (person variable). Older children in this study (5-year-olds) had an additional understanding of person variables (e.g., age) and certain strategy variables that could aid in the retrieval of items. To address the issue of discriminant validity, Wellman assessed children's understanding of variables that are both relevant and irrelevant to cognitive performance. Without this control, children could be credited with the understanding of variables that affect cognitive performance (metacognitive knowledge) in situations when they are simply stating that "everything affects everything" (Wellman, 1985, p. 197). It was found that the majority of 3-year-old children understood that, for

example, a person's hair color or weight were unrelated to cognitive performance.

Wellman (1985) noted that young children's unrealistic assessments of their memory span in Flavell et al. (1970) could be explained by the fact that memory performance is influenced not only by the number of items but also by the effort and time spent on a particular memory task. Children seemed to attribute different weights to these variables. He also stated that the degree to which children understand the interaction of cognitive variables and their concepts of the human cognitive system as a whole, including its features and functions, are still unknown.

Memory monitoring. Memory monitoring in children has often been investigated via the use of feeling-of-knowing judgments (Wellman, 1977b; Cultice, Somerville, & Wellman, 1983; Butterfield, Nelson & Peck, 1988). A feeling-of-knowing judgment is a person's insight about the status of an item in his or her memory. This judgment can be positive, which implies that one knows that he or she has some information in memory, even if the information is not accessible at the moment. Tip-of-the-tongue experiences are a subset of positive feeling-of-knowing judgments. A negative feeling-of-knowing judgment is a decision that a particular item is not stored in memory and cannot be retrieved or recognized.

Wellman (1977b) examined children's memory monitoring using feeling-of-knowing judgments and assessing tip-of-the-tongue experiences. Children from 6 to 9 years of age were presented with line drawings of concrete objects. Some of these objects were easy to name (e.g., banana), and some were less common and, therefore, harder to name (e.g., metronome). The task had three phases: naming, judgment and recognition. Children were asked to name a picture, and if they were unable to name it, they were asked whether they would be able to recognize it upon hearing it (the feeling-of-knowing judgment). The

accuracy of monitoring was measured by comparing the judgments and the results of later recognition. The results of this study suggested that children's ability to make feeling-of-knowing judgments increased with age. Kindergartners were only slightly better than chance at predicting their later recognition whereas third graders were extremely accurate.

Cultice et al. (1983) tested preschoolers' memory monitoring using the same sequence as Wellman (1977b) but with improved methodology. The naming, judgment and recognition sequence involved naming people from a set of photographs. The authors argued that proper names and real pictures, as opposed to object naming and line drawings, provided a more sensitive and sensible test for preschool-aged children because specific nominals (proper nouns) are among the first words children learn (Nelson, 1973). Moreover, even young children know that it is important to remember people's names, and they have much experience with recalling or failing to recall them. The results of this study showed that preschoolers were evaluating an item's memory status on the basis of relevant cues like familiarity. For example, some children made spontaneous comments in the naming phase such as, "she looks familiar" and "I don't know him but I have seen him". Consequently, preschoolers were rather accurate in their memory monitoring. In addition, this study demonstrated that positive feeling-of-knowing judgments lead to more retrieval efforts in school-aged children. This suggested that they were using monitoring judgments to regulate their responses.

Some researchers have argued that the original feeling-of-knowing paradigm, which utilizes absolute yes/no judgments, relies heavily on a participant's threshold for claiming to know, and this threshold may be affected by the participant's mental and chronological age (Butterfield et al., 1988). Butterfield et al. modified the original paradigm to employ relative comparison judgments. Among the participants were 6-, 10- and 18-year-olds. Relative feeling-of-knowing judgments involves a

comparison of two words. Each pair was chosen from one of the following three categories: easy and difficult words, previously correctly and incorrectly defined words, previously incorrectly and undefined words. The participant was asked to judge which word of each pair was more likely to be recognized. It was hypothesized that incorrectly defined words would yield stronger feelings-of-knowing than words for which the definition was absent. The results of this study contradicted Wellman's (1977b) notion that accuracy of memory monitoring increases with the age of a child. The data from Butterfield et al. revealed that 6-year-olds had greater feeling-of-knowing accuracy than older children. Older children tended to overestimate their recognition abilities.

Butterfield et al. (1988) suggested that even if younger children sometimes make accurate judgments about their feeling-of-knowing, they may be able to do so only under certain conditions. This may occur when making feeling-of-knowing judgments is the only activity required from them, or when they are instructed to do so. The authors indicate that these conditions need to be more accurately specified in future studies to be able to account for developmental differences in memory monitoring.

Methodological Issues

The literature suggests that one of the significant questions in the area of metacognition which needs to be addressed concerns the adequacy of tools that have been used to assess this cognitive phenomenon (Meichenbaum, Burland, Gruson & Cameron, 1985). Most metacognition studies have relied primarily on verbal data. The controversy of conscious versus unconscious components of metacognition discussed earlier poses a problem by itself in light of this fact, and there are several other issues that need to be considered as well.

First is the issue of introspective access raised by Nisbett and Wilson (1977). These authors suggested that humans might not be able to

directly access their higher order cognitive processes, and when they make an effort to recount these processes after the behavior occurred, their reports may be a reflection of post-facto rationalization rather than true introspection. In some instances these reports may be accurate, and in others not. They suggested several factors that might give rise to errors in verbal reports. One of the most prominent factors is a time gap between the occurrence of a cognitive process and the verbal report, where the accuracy of the report is inversely related to the time. Even though their original article raised a great deal of debate in the literature (for review, see Guerin & Innes, 1981; White, 1985), it nevertheless called for discretion in interpretation of results in studies that use post-performance verbal report on cognitive processes (Meichenbaum et al., 1985).

Another obstacle in assessing metacognition via the use of verbal reports stems from the fact that a significant proportion of the population under investigation may not be suitable for this form of assessment. This subset includes young children or cognitively impaired individuals. The verbal report may not be appropriate for them due to many of the following reasons: they may not comprehend the question, or be able to verbally indicate their experience, or they may simply be lacking the motivation to give a report.

Wellman (1985) noted that cross-cultural differences in the use of metaphors and mental verbs to describe internal events might also confound research findings in this area. He also stated that age-related language differences are most obvious in interview tasks, and therefore suggested that research on metacognition in young children should utilize yes-no judgments and verbal or nonverbal indication of choices rather than open-ended interview tasks.

Wellman (1985) further suggested that much of the research on concepts of cognition tapped instead the children's understanding of

mental language (e.g., words such as think, know and remember). The fact that some children fail to have the appropriate semantic knowledge is not necessarily indicative of their concepts of cognition. Johnson and Wellman (1980) assessed 4-year-old children's understanding of the following verbs: "remember", "know" and "guess". Children seemed to confuse them and they appeared oblivious to the factors that separate the mental processes described by these verbs. Only further examination of their responses revealed that these children had at least some understanding of these words, and that they could differentiate between them on the basis of their correspondence to reality. Furthermore, Johnson and Maratsos (1977) documented separate understanding of the words: "think" and "know" in 4-year-old children, but not in 3-year-olds. The failure to find differentiated understanding of these words, however, does not necessarily mean that 3-year-old children are unable to distinguish between these two mental processes. A story and question task utilized in this study to assess children's comprehension of mental verbs might have been insensitive to the abilities of the 3-year-olds because of their limited language ability.

The third problem in this area relates to the veracity of post-facto verbal reports. Information-processing theories of cognition suggest a two-process approach to thinking: automatic vs. effortful processing (Hasher & Zacks, 1979). In light of the information-processing approach, it is possible that unless the experimental task is carefully chosen, participants may engage in automatic processing when they may not be aware of their cognitions during performance (Meichenbaum et al., 1985). Borkowski (1985) suggested that conscious, and therefore reportable, metacognitive activity is likely to take place in a situation where one is engaged in a complex or novel task that involves judgment and decision making during the performance. Otherwise, participants may engage in post-facto rationalizations when offering a

verbal report on the cognitive processes that influenced their behavior (Meichenbaum et al., 1985). Therefore, studies that attempt to assess participants' cognitions and metacognitions via the use of verbal reports must choose tasks that involve conscious processes.

The previous discussion of the problems concerning verbal data, especially interviews and post-performance reports, was not meant to suggest that these techniques are worthless to metacognition research. They were specified to draw attention to the potential liabilities these techniques may possess. Meichenbaum et al. (1985) offered several additional techniques for the assessment of metacognition. Among those suggested were immediate self-report in the form of the think-aloud technique or interviews during the performance, and observation of spontaneous self-directed (private) speech. Finally, the authors proposed the inferential observation of task performance without the use of self-report as another technique that would be useful in the assessment of metacognition. Surprisingly, the employment of this behavioral measure in research has been sporadic. Even if we set aside the questions raised about the adequacy of using verbal reports per se as the tools to assess metacognition, one important problem remains. In particular, verbal reports may be less or not at all appropriate for young children and language- or cognitively-impaired individuals that have problems comprehending questions and/or verbally expressing their responses and experiences. A behavioral measure that does not employ verbal reports could improve the validity and accuracy of the assessment. In addition, it could open a window of opportunity to explore populations that were previously only partially reachable and questions that were circumscribed by a sole-focused methodology in the past.

The Uncertainty Paradigm

Creating difficulty in cognitive tasks has been used in many studies of metacognition in humans (e.g. Markman, 1977; Revelle et al., 1985; Robinson & Robinson, 1977). However, most of these studies utilized verbal tasks and required verbal responses. Recently, researchers interested in comparing the metacognitive abilities of human and nonhuman animals have used an innovative approach to the study of metacognition (Shields, Smith, & Washburn, 1997; Smith, Schull, Strote, McGee, Egnor, & Erb, 1995; Smith, Shields, Allendoerfer, & Washburn, 1998; Smith, Shields, Schull, & Washburn, 1997). The researchers in this area did not have the luxury of all of their participants being verbal. Therefore they employed an uncertainty paradigm in which simple perceptual and memory tasks were used. These were designed to create difficulty for the participants. Crucially, participants were also provided with a concrete response that allowed them to deal with the difficulty and uncertainty. This response was termed the uncertain or "escape" response because it allowed participants to abandon trials of their choosing.

Smith et al. (1998) used the uncertainty paradigm to test memory monitoring in human adults and nonhuman primates. First, a relatively simple, nonverbal memory task was required so that rhesus monkeys could participate. Second, some trials had to be made more difficult or uncertain than others. Third, an uncertain response had to be made available. The researchers then determined whether uncertain responses occurred more for difficult than for less difficult trials.

The task that Smith et al. (1998) chose was a serial probe recognition task. In a classic serial probe recognition (SPR) task, a list of pictorial items is presented to participants. Subsequently, a probe item is presented. Participants are given a binary response option and instructed to respond "there" if the probe was from the previously

presented list or "not there" if the probe was not presented in the list. Smith et al. elegantly innovated the SPR task by expanding the binary response pool by adding a third response, the uncertain response. This third response allowed participants to abandon the current trial in favor of an easier trial.

To fulfill the requirement of differential difficulty among trials, Smith et al. (1998) took advantage of the well-known serial position effect, one of the most robust of memory phenomena. Ebbinghaus (1885/1913) labeled the improved memory for items at the beginning of the list a primacy effect, and the augmented memory for the final items a recency effect. This U-shaped serial position curve has been documented in memory studies utilizing recall and recognition tests, with lists of items consisting of words, non-sense syllables and pictures (for a review, see Crowder, 1976). The serial position effect has been found in humans as well as in several nonhuman animal species (e.g., Glanzer & Cunitz, 1966; Sands & Wright, 1980; Roberts & Kraemer, 1981; Wright, Santiago, Sands, Kendrick & Cook, 1985).

Smith et al. (1998) suggested that in the memory task utilized in their experiment, participants would experience an increased amount of uncertainty for the middle section of the memory items relative to the uncertainty experienced for the initial and final items. Therefore, participants should use the uncertain response more for the mid-list items than for the initial and final items. Indeed, these results were obtained. The participants confronted the memory task in an adaptive manner. They appeared to engage in memory monitoring and employed the escape response when they encountered the most indeterminate memory traces.

In addition, Smith et al. (1998) proposed that the overall performance on trials in which the escape response is available should be superior to the performance on trials where the escape response is

not allowed. This means that when the escape response is available, participants can either escape, if they feel uncertain, or they can volunteer a primary response ("there" or "not there"), if they are more confident of the correct answer. If the participants monitor their memory adaptively, the escape response allows them to avoid trials for which they do not know the correct answer. Under these circumstances, if participants actually volunteer a primary response, their error rate should be relatively low. These "voluntary try" trials should yield a higher proportion of correct responses than the "forced try" trials, in which the escape response is unavailable and participants are forced to make a primary response even for the probes about which they feel uncertain. Again, the results of their study supported this hypothesis.

The procedure employed by Smith et al. (1998) has been used to test memory monitoring in human adults and nonhuman primates. It has never been used to test cognitive monitoring in children. Given the aforementioned problems in assessing metacognition, the simplicity of this task and its nonverbal nature could make it a feasible tool in testing these cognitive mechanisms in children.

Support for Using the SPR Task

Although the SPR task has not been used to examine memory monitoring in children, other memory phenomena have been investigated via the use of similar techniques. A probe-type pictorial memory task has been used to investigate short-term memory processes in children of various ages (Atkinson, Hansen & Bernbach, 1964; Hansen, 1965; Calfee, Hetherington & Waltzer, 1966) as well as in children with mental retardation (Ellis & Munger, 1966) in the past. The serial position effect was documented in these studies of memory.

The recognition abilities in children of various ages have been documented using several different techniques. One of these procedures involves presenting a list of items to participants while asking them to

indicate whether or not an item is a repeat. Corsini, Jacobus & Leonard (1969) tested preschool children's recognition skills for words and pictures using this procedure and found a superior recognition for pictures. Evidence of good recognition memory for pictures in preschool children was subsequently supported by many researchers (e.g. Brown and Scott, 1971; Perlmutter & Myers, 1974, 1976). Brown and Campione (1972) noted that the ability to detect repeating pictures is well established in preschool children, and their recognition of distinct pictorial stimuli is similar to that which has been reported for adults. The research suggested no reason to expect hindrance in using a serial probe recognition task in order to assess memory monitoring in preschool children.

In addition, Butterfield et al. (1988) in their study of the developmental aspects of memory monitoring emphasized the need for making a distinction between a person's potential monitoring ability and the utilization of the products from the person's monitoring. However, the most frequently used paradigm for assessing memory monitoring, the feeling-of-knowing judgment, does not allow for this distinction, because it taps the former but it leaves out the latter. The feeling-of-knowing judgment is a person's insight about the status of an item in his or her memory. Young children may discriminate information from feeling-of-knowing judgments in their memories but it is unclear whether they can spontaneously use it in a proper way (Butterfield et al., 1988). Clearly, a behavioral measure assessing memory monitoring is needed.

Finally, the cognitive literature has made an insistent call for improved assessment of cognitive monitoring in children, such as the use of simple, well-defined cognitive tasks, which could tap children's cognitive monitoring without introducing confounding variables built into the assessment tools utilized in past studies. Developmental

psychologists recorded some of the metacognitive limitations of younger children (Baker, 1994). On the other hand, one should keep on mind that science can be no better than its methods. Therefore, among the most important responsibilities one has as a scientist is to continually search for improved ways to explore the phenomena in question.

This study employed the inferential observation of task performance as a behavioral measure to assess memory monitoring in five-year old children. The uncertainty paradigm of Smith et al. (1998) served as a procedural guideline to induce and consider memory monitoring. The indicated procedure was used to minimize the problems of discriminant validation as well as the potential liabilities of utilizing verbal reports in the assessment of metacognition.

It was hypothesized that five-year-olds would behaviorally demonstrate adept memory monitoring via the use of the uncertain response when the most indeterminate memory traces were probed. Specifically, their use of the uncertain response would be an invert reflection of the serial position curve of their memory performance. Moreover, when their performance is compared to a group that was denied the escape response, the performance of the former would be superior. This reduction in error rate would demonstrate an adaptive use of the uncertainty response and memory monitoring.

Method

Participants

Participants of this experiment were 43 4.5- to 5.5-year-old children (mean age = 5.09, SD = 0.33) from the Missoula, Montana, community. Recruitment of the participants was conducted via information sheets distributed to the parents by the investigator, her research assistant and the directors of day-care institutions. The parents were asked to make an appointment either in person or by placing a phone call to the investigator if they were interested in the study. Participants

were randomly assigned to the following two groups: experimental (21 children) and control (19 children). The data obtained from three children were excluded from the analysis because of their inability and/or lack of motivation to follow the instructions.

Apparatus

An IBM compatible computer was used to present pictorial stimuli on a monitor in both the pilot study and the actual experiment. An EZscreen™ touch screen add-on attached to a 15-inch color monitor recorded the responses. Participants were seated on a height-adjustable chair at arm's length away from the monitor.

Procedure

Prior to the experiment, an informed consent form was filled out by a parent or legal guardian of each child.

The experimental procedure was analogous to the one used in Smith et al. (1998), Experiments 1 and 3. Each trial consisted of two phases. The first phase involved the presentation of a list of to-be-remembered items, and the second phase consisted of the presentation of response choices. The number of choices depended on whether a participant was assigned to the control or experimental group. The control group was given a classic (binary-response) serial probe recognition task. The experimental group was allowed an additional choice--the escape or uncertainty response. The participants were assigned to the control and experimental groups randomly.

Stimulus presentation phase. On each trial, participants made a voluntary trial-initiation response by touching a white circle 1 cm in diameter positioned in the upper left corner of the screen. Below the circle, there was a number that indicated the number of the trial. Participants were then presented with a list of pictures chosen from among 81 digitized color drawings of common objects such as an apple, a

tiger, and a camera. Prior to this study, four five-year-olds were asked to name 101 drawings that had been used by Smith et al. (1998). The selected 81 drawings were those that all of the children were able to name. Each picture was presented sequentially in the upper left of the screen. The list length started at 6 items, and increased by one with each new trial. After four trials the list length reached 10 items and no further increments to the list length were made. The first four trials served as a transition from the practice session, and were excluded from later analysis. The pictorial stimuli for each trial were selected randomly. Repetitions within a trial's list were not allowed but no other limitations on the frequency of repetitions were imposed. Each item was displayed for approximately 900 ms, which was the time required to draw the picture on the monitor. The presentation of the next item followed immediately. In the case of the final list item, a retention interval of 1 s followed the presentation of the item.

Response phase. The response phase differed between the control and experimental groups. For the control group, a probe picture was displayed in the upper left of the screen. A red icon with a large, white letter X was displayed in the upper right of the screen. The red icon was approximately the same size as the probe picture. If the probe was from among the pictures presented on the preceding list, a correct response was to touch the probe picture. If the probe was absent from the list, a correct choice was to touch the letter X. A correct response earned a computer-generated reward sound, verbal praise and a marble dispensed by the experimenter. An incorrect response resulted in a penalty sound (a brief buzz of approximately 500 ms) and no marble reward. These were the only response options presented to the control group. Thus, the control group performed the classical serial probe recognition task.

The experimental group followed the same procedure, except that it was also offered an escape choice--the uncertain response. On each trial a yellow icon with a black question mark (?) in the middle was illuminated in the lower center of the screen along with the probe picture and the letter X in their previously specified locations. All of the icons were approximately the same in size (3 x 3.5 inches). Choosing the uncertain response cleared the current trial. A new trial, in which only one response was illuminated, followed. If the probe was from the list, the probe picture appeared on the screen. If it was not, the X was displayed. The reward was secured by touching the displayed object.

In order to prevent the overuse of the uncertainty choice, its cost was actively modified. The program calculated the ratio of recent escapes to primary responses. If (recent escapes - recent primaries) exceeded four, the use of the uncertain response froze the touchscreen for a period of time. This period was determined by the formula (recent escapes - recent primary responses)² x 50 ms. Therefore, the overuse of the uncertain response resulted in the delayed opportunity to give a response that would be registered by the touchscreen and thus in the delayed reward. The cost of overusing the uncertain response was designed to discourage participants from employing it when they knew the correct answer, saving it for the instances in which their memory trace was more indeterminate. The experimenter gave the children a verbal explanation of this circumstance, stating the "the computer is tired from so many questions, because we have been asking the computer for help too many times".

The computer was programmed so that 40% of the trials were those in which the probe picture was absent from the list ("not there" trials), and the remaining 60% were "there" trials. This trial distribution was chosen to maximize the number of trials per serial position but at the same time not to prompt children to use the probe as

a default response. The computer program was set to present the trials using the following rules: for every 10 trials, six were "there" and four were "not there" trials; for every ten "there" trials, there was one probe per serial position.

The experiment lasted approximately 40 minutes, and each child completed an average of 45 experimental trials in addition to the average of 5 practice trials. At the beginning of the experiment, each participant heard the instructions (see Appendix) told by the experimenter. Each participant was given several trials to practice responding using the touchscreen. The practice trials differed from the experimental trials in their list length. The list length for the practice trials started at 4 items, and increased by one with each new trial. During this time, the researcher had an opportunity to check for children's correct understanding of the instructions. The children were asked whether or not they saw the probe picture, and what would be the appropriate response to give in that situation. If they responded correctly, the practice trial session was terminated, and the experimental session was started.

At the end of the experiment, the children were given positive verbal feedback. The experimenter reassured them of their good performance on the task. Each child received on average 8 stickers of his or her choice from a motivational sticker collection in exchange for the marbles, irrespective of marbles gained. In addition, children who participated in the experiment (or their parents) received five dollars in gift certificates to Dairy Queen or cash, depending on their choice.

Pilot Study

There was a pilot study conducted prior to the actual experiment. Eleven 5-year-olds (mean age = 5.07, SD=0.53) participated in this study. One child's results were excluded from the analysis because of his inability and/or lack of motivation to attend to the presentation of

the pictorial stimuli. Children from the pilot study were not eligible for participation in the later study.

Children in the pilot study performed the same experimental task as described above, but the task parameters were manipulated across participants. These calibration trials were designed to determine parameters that would yield the appropriate level of difficulty for the experimental task. Careful examination of children's performance under different conditions allowed us to determine more appropriate length of the interstimulus interval, stimulus presentation duration, as well as the list length, and the pattern of the practice trials. In addition, the pilot study indicated the necessity of implementing the trial-initiation response that had allowed children to focus their attention to the appropriate area of the display screen immediately prior to the stimulus presentation.

Results

Primary Memory Performance

Children's primary memory performance was considered for a list length of 10 items. The first four trials for each child with list lengths of 6, 7, 8 and 9 items were aimed only for training purposes and were excluded from the analysis. Table 1 summarizes the percentage correct for each serial position, all the trials combined and all "not there" trials for the experimental and control groups.

Insert Table 1 about here

The possibility of temporal effects, such as fatigue, was explored by dividing each child's data into two equal halves. The first and second halves of the children's trials were analyzed separately. Table 2 summarizes the percentage of correct for this split-half data.

Insert Table 2 about here

Memory Monitoring and the Use of the Uncertain Response

In the experimental group, children's performance was further evaluated on the basis of their use of the uncertain (escape) response. Figure 1A illustrates their escape behavior in relation to their primary memory performance. For comparison, the primary memory performance of the control group, which was not allowed the escape response, is depicted in Figure 1B.

Insert Figure 1 about here

Split-half analysis was again conducted to evaluate the potential temporal effects on the data. Figure 2 depicts this divided performance for the experimental and control groups. Figures 2A and 2B illustrate the performance of the experimental group on the first and second halves of trials, respectively. Correspondingly, figures 2C and 2D depict the performance of the control group on the first and second halves of trials, respectively.

Insert Figure 2 about here

Signal Detection Analysis of Performance

The data are particularly well suited to a signal detection analysis, a technique commonly used in studies on perception and cognition. In signal detection theory, as described by MacMillan and Creelman (1991), performance is analyzed according to the degree to which the participants' responses reflect the stimuli and the degree to which participants show bias in their responding.

Sensitivity. Sensitivity refers to the participant's ability to discriminate between stimuli. In this context, sensitivity was measured by the degree to which "there" and "not there" responses mirrored "there" and "not there" items. If the probe was from among the pictures presented on the list ("there" item), a correct response was to touch the probe picture ("there" response). In signal-detection theory terminology, correctly recognizing the "there" item is termed a hit; while failing to recognize it (touching the X) is a miss. If the probe was absent from the list ("not there" item), a correct choice was to touch the letter X ("not there" response). Correctly responding "not there" is termed a correct rejection. Erroneously recognizing an item as "there" is a false alarm. The data can be summarized in a 2x2 table displaying either the frequencies of these four events occurring or their response rates calculated as the proportion of trials the participant made a particular response. Table 3 shows the frequencies and proportions of the correct responses (hits and correct rejections) as well as the incorrect responses (misses and false alarms) for the experimental and control groups.

Insert Table 3 about here

A sensitivity measure can be obtained from the discrepancy between a hit rate and a false alarm rate. Signal Detection Theory offers a sensitivity measure called d' (d-prime). Computing d-prime requires first a z-transformation, which converts a hit and false alarm rate to standard deviation units or z-scores, and then d-prime is obtained as a difference between these transformed hit and false alarm rates. Hence, the resulting formula for d-prime is:

$$d' = z(H) - z(F).$$

First, the overall d-primes for the experimental and control groups were calculated. The d-prime for the experimental group was 1.08 and the d-prime for the control group was 0.73. Thus, the experimental group showed an increase of 0.35 standard deviation units in sensitivity as compared to the control group.

Second, the individual d-primes were computed for each participant in both groups. A directional, two-independent samples t -test was conducted on these d-primes to compare the sensitivity of the experimental and control groups ($t(38)=1.585$, $p=0.061$). These findings suggest that there was no difference at $\alpha=0.05$ level between the means of d-primes for the experimental and control groups.

Third, each data file was again divided into two equal halves to check for possible temporal effects. Two d-primes, one for the first half, the other for the second half of the trials were computed for each child. A directional two-independent samples t -test was performed to compare the differences in sensitivity between the experimental and control groups on the first half of the trials ($t(38)=2.064$, $p=0.023$). The same statistical technique was executed for the d-primes from the second half of the trials ($t(38)=0.7$, $p=0.244$). These results indicate that the mean d-prime for the experimental group was significantly higher than that of the control group for the first half of the trials. However, this difference was eliminated for the second half of the trials. Figure 3 illustrates the mean d-primes for the experimental and control groups in the first and the second halves of trials.

Insert Figure 3 about here

Response bias. Response bias refers to the inclination to favor one response over another. The most commonly used bias measure for

signal detection analysis is called criterion location (c), and is defined as $c = -0.5[z(H) + z(F)]$.

In the context of the presented experimental scenario this means that a positive bias is a tendency to favor the "not there" response (X). Consequently, a negative bias is a tendency to prefer "there" response (probe).

An overall response bias as measured by the criterion location (c) was computed for the experimental and control group. The control group exhibited a relatively strong negative bias ($c = -0.298$) preferring "there" over the "not there" response. This was more than double that of the bias of the experimental group ($c = -0.131$).

Discussion

Both experimental and control groups exhibited serial position effects in their primary memory performance (see Figures 1A and 1B). However, the escape responses for the experimental group did not mirror their primary memory performance (Figure 1A). It is possible that the variability in memory abilities among participants contributed to the lack of positive findings. Although their primary memory performance mimicked the well-established findings of serial position effects, the feelings of uncertainty may be more subjective in this age group. These feelings of uncertainty may have differed in content and degree from one child to another and thus may have been more susceptible to the averaging effects present in the graphs of overall performance, which are insensitive to inter-individual differences. Although studies with human adults have also used the between-subject design and evaluated the overall performance of the whole group, younger children may exhibit greater inter-individual fluctuations in performance. This can occur because their cognitive abilities including memory monitoring are subject to change during the course of development. Unfortunately, it was not feasible to obtain a clear graphical description of each child's

performance on different serial position trials due to the relatively small number of trials for each child.

Additionally, the overall performance was evaluated on a temporal basis for possible practice effects and/or effects of fatigue. Given the inconclusive results, it was suspected that children's performance might have changed during the course of the experiment. Therefore, the data from each child was divided into two equal halves. The overall performance on the first half of the trials was evaluated separately from the data on the second half. Alas, this division, which resulted in a smaller number of trials evaluated in each group, led to an even greater variability in the results rendering the detection of stable trends via the means of graphical analysis unattainable.

It is possible that, in addition to the differences in ability, the relatedness of some stimuli and their repetitions affected primary memory performance and memory monitoring accuracy. Except for repetitions within a trial's list that were not allowed, no other restrictions on the frequency of repetitions were imposed. Therefore, it is plausible to assume that participants, who viewed a particular picture in a trial's list at an easier serial position earlier on in the game, could have carried over their memory of this picture to the latter trials. In the latter trials this picture was tested as a probe from a more difficult serial position. This more difficult-to-remember serial position that was hypothesized to yield a more indeterminate memory trace, although remembered for wrong reasons, nevertheless produced a correct response. Hence, participants' performance might have been influenced by these repetitions, which led to an obscured picture of memory monitoring occurrence.

The visual similarity of some stimuli could have also obfuscated participants' memory performance. Children may have confused one picture with another if these shared several visual features (e.g. a picture of

a lioness and a picture of a tiger, or a printer and a typewriter). In addition, the visual attractiveness or a personal significance of some stimuli could have obscured memory performance as well. Although the presentation of stimuli was very brief, and for the most part children did not have the time to verbally label or comment on the stimuli, it has happened that a child, delighted by a particular picture, labeled the picture during the presentation phase. If this picture then came up as a probe, the memory of it was enhanced by the child's verbalization and, regardless of the serial position of this picture, the child gave a correct answer.

Furthermore, as Kelemen, Frost and Weaver (2000) suggested, guessing generates noise that can distort metacognitive accuracy in recognition predictions. Thus, participants, who do in fact monitor their memory accurately, might exhibit ambiguous metacognitive performance because they guessed on the items they did not know. Consequently, correct guessing could have contributed to the added variability within our data and prevented us from finding the escape curve that would mirror participants' primary memory performance.

The aforementioned reasons have led to a search for a measure of cognitive performance that could evaluate the overall performance of both of the groups, as well as that of the individual participants, without taking into consideration the serial positions of the probes. It was suggested that signal-detection analysis (SDA) could handle this task. There is a particular advantage to using SDA as the means for evaluating memory performance. SDA offers a measure of memory sensitivity that is independent of the participants' bias to favor one response over another. The mechanics of SDA were explained in the Results section.

Only the hit rate was explicitly utilized in the initial graphical analysis, which could have lead to a rather inaccurate picture. With

only one number (proportion of hits) at hand, one would miss an important piece of information. Obviously, a participant who yields a hit rate of 0.85 and a false alarm rate of 0.3 is more sensitive than a participant who generates a hit rate of 0.85 but also a false alarm rate of 0.85. Clearly, the second case indicates that the participant favored the "there" response with no correspondence to the actual stimulus. Hence, the proportion of correct responses as a single indicator of performance is appropriate only when there is no tendency to prefer one response over another. In this study, the proportion of "there" and "not there" trials was not equal, but was instead set on 60% and 40%, respectively. Therefore, the signal-detection analysis offered a more appropriate measure of performance, given the experimenter-implemented bias, in addition to the bias potentially used by our participants:

The comparison of individual d-primes was not statistically significant, suggesting no difference between the experimental and control groups with regard to sensitivity.

However, the results indicate that the control group had a tendency to prefer the "there" response over the "not there" response. This was documented by their negative response bias, which was more than double the bias of the experimental group. It is possible that participants in the control group used the "there" response as a default when unsure. Paradoxically, this in turn could have contributed to the higher overall percentage of correct responses because the distribution of "there" and "not there" trials was set on 60% and 40%, respectively. In other words, this means that even if the participant was completely oblivious to the stimuli and used the "there" response on all trials indiscriminately, this participant would still yield a rate of 60% correct responses.

Perhaps our failure to find the significant overall improvement in memory performance for the experimental group was in part caused by the

fact that the control group used the "there" response as a default in situations when unsure. Consequently, this guessing, which was correct the majority of the time due to the trial distribution, inadvertently improved their performance.

Still, the subsequent analyses of the first and the second halves of children's trials indicated that the sensitivity in discrimination between stimuli was different for the experimental and control groups. The participants in the experimental group were significantly more sensitive in discrimination between stimuli than children in the control group for the first half of their trials. This sensitivity, however, changed during the second half of the experiment when there were no significant differences between the experimental and control groups. These findings suggest that the performance of the experimental group was superior to that of the control group during the first half of the experiment. During the second half of the experiment, fatigue could have played a role in children's ability to discriminate between stimuli annulling the advantage of the experimental group. The fatigue could have resulted from the increased demands on children in the experimental group, who were required to perform an additional task, namely, to explicitly monitor their memory in terms of using the uncertain response. Children in the control group did not have this additional demand on their resources, and in fact their mean performance was marked by a slight increase in sensitivity for the second half of the trials (see Figure 3). Children in the control group actually might have benefited from additional practice as the trials progressed.

Although the effects of fatigue and practice were not the subjects of inquiry in the present study, the possibility of these two factors influencing the results is worth mentioning and certainly worthy of exploration in future research endeavors. In the light of discussion recently revisited in the metacognition literature pertaining to the

issue of conscious versus unconscious metacognitive processing, the potential effect of fatigue on the ability to monitor memory processing is very interesting. Hasher and Zacks (1979) suggested that the processes underlying cognition can be either automatic or effortful. Effortful processing requires conscious attention, but automatic processing makes minimal demands on attentional resources. If memory monitoring is indeed subject to the effects of fatigue, then that would suggest that this type of metacognition requires conscious processing. It would be of course premature to make any conclusions pertaining to this issue, but this idea may serve as a pointer for future research.

Several comments and improvements are suggested for future investigations. A larger sample size would increase the power of the statistical analyses and the potential effect size. The second recommendation relates to the innovation of the experimental design. A within-subject, repeated measures design that would yield a higher number of trials per participant seems a logical next step in this line of research. If the participants could be tested on several occasions, serving as their own controls, the noise created by large inter-individual differences would be dramatically reduced. Additionally, it would allow clearer answers for some questions generated in this study such as: Does the experimental task that offers the escape/uncertain response make additional demands on attentional resources or is the performance of the control task influenced by practice?

The experimental task itself could also be improved. For example, imposing more severe restrictions on the frequency of stimuli repetitions, and changing the trial distribution ratio could help to obtain a clearer picture of primary memory and memory monitoring performance. Substituting pictures of familiar objects for pictorial stimuli with distinct geometrical shapes could decrease problems with verbal labeling, as well as the possible personal significance of

stimuli. On the other hand, this change could decrease the attractiveness of the experimental task itself and children's interest in playing the game would be compromised.

Having learned about some of the difficulties in using this experimental paradigm for testing five-year-old children, it would be interesting to investigate younger children with an improved design. A developmental comparison of memory monitoring abilities across several different ages would not only be exciting but also very instructive in the field of metacognitive research.

References

- Atkinson, R. C., Hansen, D. N., & Bernbach, H. A. (1964). Short-term memory with young children. Psychonomic Science, 1, 255-256.
- Baker, L. (1994). Fostering metacognitive development. Advances in Child Development, 25, 201-239.
- Brown, A. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology. (pp.77-165). Hillsdale, NJ: Erlbaum.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding. (pp.65-116). London: Lawrence Erlbaum.
- Brown, A.L., Campione, J. C. (1972). Recognition memory for perceptually similar pictures in preschool children. Journal of Experimental Psychology, 95(1), 55-62.
- Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R. Siegler (Ed.), Children's thinking: What develops? (pp. 3-35). Hillsdale, NJ: Erlbaum.
- Brown, A. L., Scott, M. S. (1971). Recognition memory for pictures in preschool children. Journal of Experimental Child Psychology, 11, 401-412.
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling-of-knowing. Developmental Psychology, 24, 5, 654-663.
- Calfee, R. C., Hetherington, E. M., Waltzer, P. (1966). Short-term memory in children as a function of display size. Psychonomic Science, 4, 153-154.
- Cavanaugh, J. C., & Perlmutter M. (1982). Metamemory: A critical examination. Child Development, 53, 11-28.

Corsini, D. A., Jacobus K. A. & Leonard, S. D. (1969). Recognition memory of preschool children for pictures and words. Psychonomic Science, 16(4), 192-193.

Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling-of-knowing judgements. Child Development, 54, 6, 1480-1486.

Ebbinghaus, H. (1885/1913). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius, Trans.). New York: Columbia University. (Original work published 1885)

Ellis, N. R. & Munger, M. (1966). Short-term memory in normal children and mental retardates. Psychonomic Science, 6(8), 381-382.

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of. Human Development, 14, 272-278.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. American Psychologist, 34, 10, 906-911.

Flavell, J. H. (1981). Cognitive monitoring. In W. P. Dickson (Ed.), Children's oral communication skills. New York: Academic Press.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding. (pp.21-29). London: Lawrence Erlbaum.

Flavell, J. H., Friedrichs, A. G., Hoyt, J. D. (1970). Developmental changes in memorization processes. Cognitive Psychology, 1, 324-340.

Garner, R. (1987). Metacognition and reading comprehension. Norwood, NJ: Ablex.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. Journal of Verbal Learning and Verbal Behavior, 5, 352-360.

Gombert, J. E. (1993). Metacognition, metalanguage, metapragmatics. International Journal of Psychology, 28, 5, 571-580.

Guerin, B., & Innes, J. M. (1981). Awareness of cognitive processes: Replications and revisions. Journal of General Psychology, 104, 2, 173-189.

Hansen, D. N. (1965). Short-term memory and presentation rates with young children. Psychonomic Science, 3, 253-254.

Hart, J. T. (1965). Memory and feeling of knowing experience. Journal of Educational Psychology, 56, 208-216.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 108, 356-388.

Johnson, C. N., & Maratsos, M. P. (1977). Early comprehension of mental verbs: Think and know. Child Development, 48, 1743-1747.

Johnson, C. N., & Wellman, H. M. (1980). Children's developing understanding of mental verbs: Remember, know, and guess. Child Development, 51, 1095-1102.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. Psychological Review, 88, 67-85.

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. Memory and Cognition, 28(1), 92-107.

Kuhn, D. (1989). Children and adults as intuitive scientists. Psychological Review, 96, 674-689.

MacMillan, N. A., & Creelman, C. D. (1991). Detection Theory: A user's guide. New York: Cambridge University Press.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. Child Development, 48, 986-992.

Meichenbaum, D., Burland, S., Gruson, L., & Cameron, R. (1985). Metacognitive assessment. In S. R. Yussen (Ed.), The growth of reflection in children (pp.3-35). New York: Academic Press.

- Nelson (1973). Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development, 38.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 3, 231-259.
- Perlmutter, M. & Myers, N. A. (1974). Recognition memory development in two- to four-year-olds. Developmental Psychology, 10(3), 447-450.
- Perlmutter, M. & Myers, N. A. (1976). Recognition memory in preschool children. Developmental Psychology, 12(3), 271-272.
- Revelle, G. L., Wellman, H. M., & Karabenick, J. D. (1985). Comprehension monitoring in preschool children. Child Development, 56, 654-663.
- Roberts, W. A., & Kraemer, P. J. (1981). Recognition memory for lists of visual stimuli in monkeys and humans. Animal Learning and Behavior, 9, 587-594.
- Robinson, E. J., & Robinson, W. P. (1977). Development in the understanding of causes of success and failure in verbal communication. Cognition, 5, 363-378.
- Sands, S. F., Wright, A. A. (1980). Serial probe recognition by a rhesus monkey and a human with 10- and 20-item lists. Journal of Experimental Psychology: Animal Behavior Processes, 6, 386-396.
- Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. Cognitive Psychology, 10, 271-301.
- Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain responses by humans and rhesus monkeys (Macaca mulatta) in a psychophysical same-different task. Journal of Experimental Psychology: General, 126, 147-164.

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., Erb, L. (1995). The uncertain response in the bottlenosed dolphin (Tursiops truncatus). Journal of Experimental Psychology: General, 124, 391-408.

Smith, J. D., Shields, W. E., Allendoerfer, K. R., & Washburn, D. A. (1998). Memory monitoring by animals and humans. Journal of Experimental Psychology: General, 127, 227-250.

Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. Cognition, 62, 75-97.

Wellman, H. M. (1977a). Preschoolers understanding of memory relevant variables. Child Development, 48, 1720-1723.

Wellman, H. M. (1977b). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. Child Development, 48, 13-21.

Wellman, H. M. (1983). Metamemory revisited. In M. T. Chi (Ed.), Contributions to human development: Vol 9. (pp. 31-51). Basel: S. Karger.

Wellman, H. M. (1985). The child's theory of mind: The development of conceptions of cognition. In S. R. Yussen (Ed.), The growth of reflection in children (pp. 169-206). New York: Academic Press.

White, P. A. (1985). The awareness issue and memorial influences upon accuracy of verbal report: A re-examination of some data. Psychological Reports, 57, 1, 312-312.

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. Science, 229, 287-289.

Appendix

Experimental Group Instructions

We are going to play a game now. First, you have to touch a circle that will come up on the screen. The computer will then show you a list of eight pictures. You watch those pictures carefully. Then, the computer will show you one picture, and two signs [the experimenter shows the child two cards, one with the letter X, the other with the question mark ?]. If this picture was in the list that the computer showed you, let the computer know by touching the picture. If it wasn't in the list, touch the letter X. If you don't know, touch the sign ?. If you touch the sign ?, the computer will show you a correct answer. Then touch that answer.

If you were right, the computer will make a beep sound and I will give you a marble. We will play this game many times. Try to get as many marbles as you can. At the end of our playing, you can exchange the marbles you won for a toy you choose. We have some toys here.

If you were wrong, the computer will make a buzz, and I will not give you a marble. After that we will start again and play a new game. You start a new game by touching the circle that will appear on the screen.

Now, let me show you how it works. If you don't understand, just ask me, and I will show you again. Then, you will try it three times, and after that, you can start playing.

Control Group Instructions

We are going to play a game now. First, you have to touch a circle that will come up on the screen. The computer will then show you a list of eight pictures. You watch those pictures carefully. Then, the computer will show you one picture, and one sign [the experimenter shows the child one card with the letter X]. If this picture was in the list

the computer showed you, let the computer know by touching the picture. If it wasn't in the list, touch the letter X.

If you were right, the computer will make a beep sound and I will give you a marble. We will play this game many times. Try to get as many marbles as you can. At the end of our playing, you can exchange the marbles you won for a toy you choose. We have some toys here.

If you were wrong, the computer will make a buzz, and I will not give you a marble. After that we will start again and play a new game. You start a new game by touching a circle that will appear on the screen.

Now, let me show you how it works. If you don't understand, just ask me, and I will show you again. Then, you will try it three times, and after that, you can start playing.

Table 1

Percentage of Correct Responses for List Length 10

Trial Type	Groups	
	Experimental	Control
SP1	76	75
SP2	78	70
SP3	68	53
SP4	66	70
SP5	74	76
SP6	71	80
SP7	73	74
SP8	77	93
SP9	80	76
SP10	85	77
All T	75	75
All NT	66	53
Overall	71	66

Note: SP = serial position; All T = all "there" trials combined; All NT = all "not there" trials combined

Table 2

Percentage of Correct Responses for the Split-Half Data

Trial Type	<u>Experimental Group</u>		<u>Control Group</u>	
	1st half	2nd half	1st half	2nd half
SP1	88	65	72	78
SP2	71	84	74	65
SP3	75	64	67	37
SP4	73	56	58	82
SP5	74	75	67	84
SP6	77	65	76	86
SP7	69	76	60	87
SP8	87	65	95	92
SP9	83	76	77	75
SP10	86	83	83	71
ALL T	79	71	73	76
ALL NT	67	64	56	49
OVERALL	74	68	66	65

Note: SP=serial position; All T=all "there" trials combined; All NT=all "not there" trials combined

Table 3

Signal-Detection Distribution of Responses

Response Type	GROUP	
	Experimental	Control
Hits	341	283
False alarms	145	83
Misses	116	95
Correct rejections	161	160

Figure Captions

Figure 1A. Serial probe recognition performance for the experimental group. Percentage of correct responses and escaped trials (when applicable) is shown. NT = "not there" trials in which the probe had not been in the preceding list.

Figure 1B. Serial probe recognition performance for the control group. Percentage of correct responses and escaped trials (when applicable) is shown. NT = "not there" trials in which the probe had not been in the preceding list.

Figure 2A. Serial probe recognition performance on the first half of the trials for the experimental group. NT = "not there" trials in which the probe had not been in the preceding list.

Figure 2B. Serial probe recognition performance on the second half of the trials for the experimental group. NT = "not there" trials in which the probe had not been in the preceding list.

Figure 2C. Serial probe recognition performance on the first half of the trials for the control group. NT = "not there" trials in which the probe had not been in the preceding list.

Figure 2D. Serial probe recognition performance on the second of the trials for the control group. NT = "not there" trials in which the probe had not been in the preceding list.

Figure 3. Change in mean sensitivity (d' -primes) for the first and second halves of sessions for the experimental and control groups.

Figure 1A.

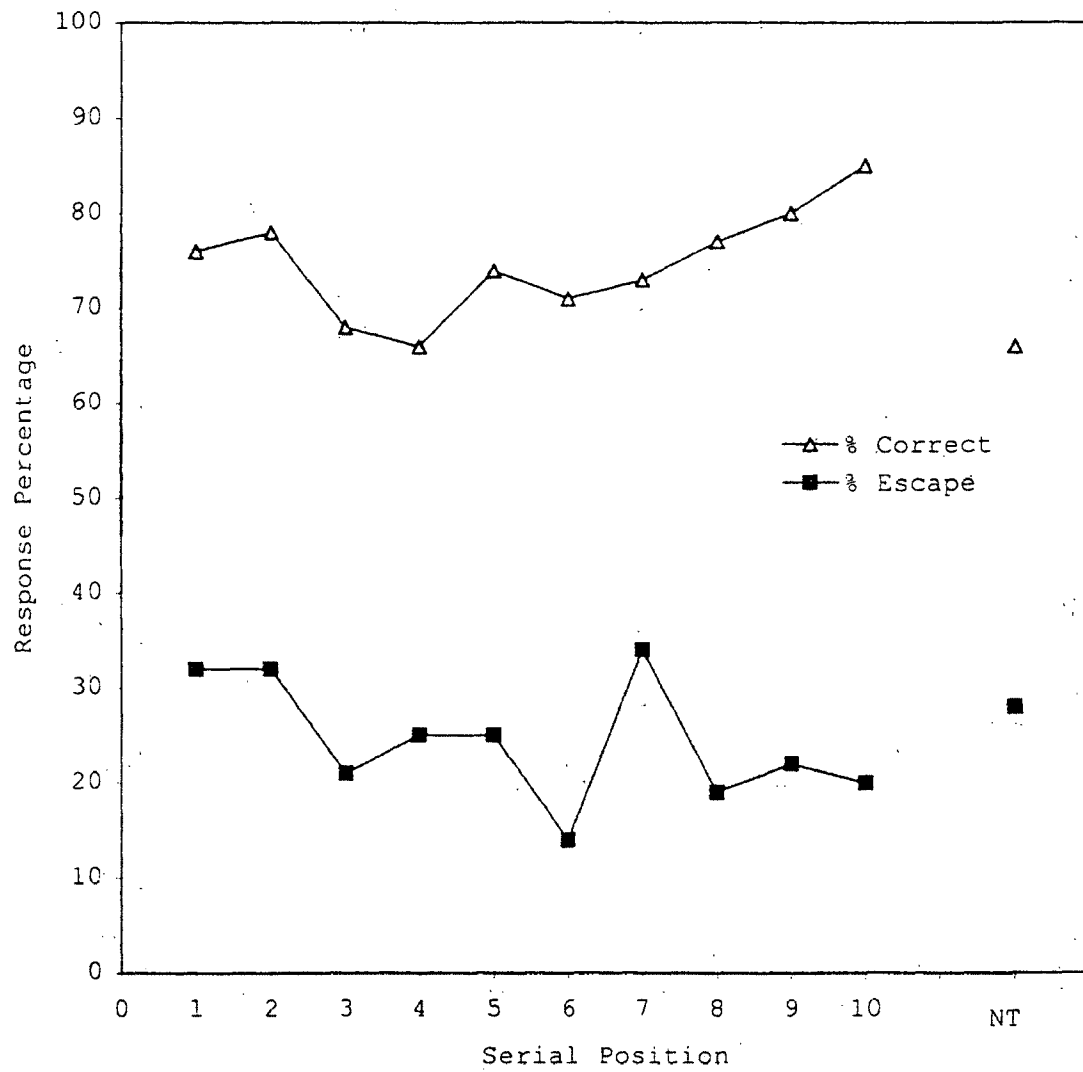


Figure 1B.

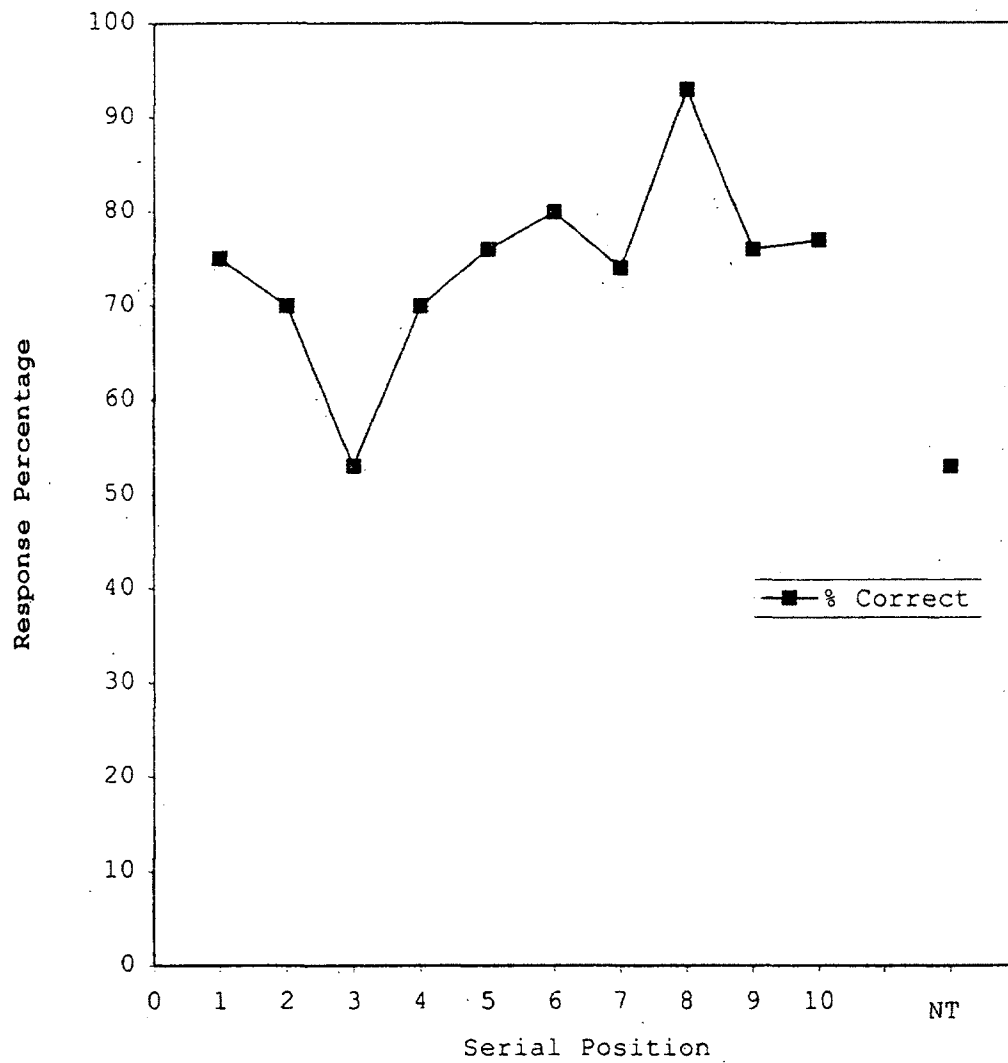


Figure 2A.

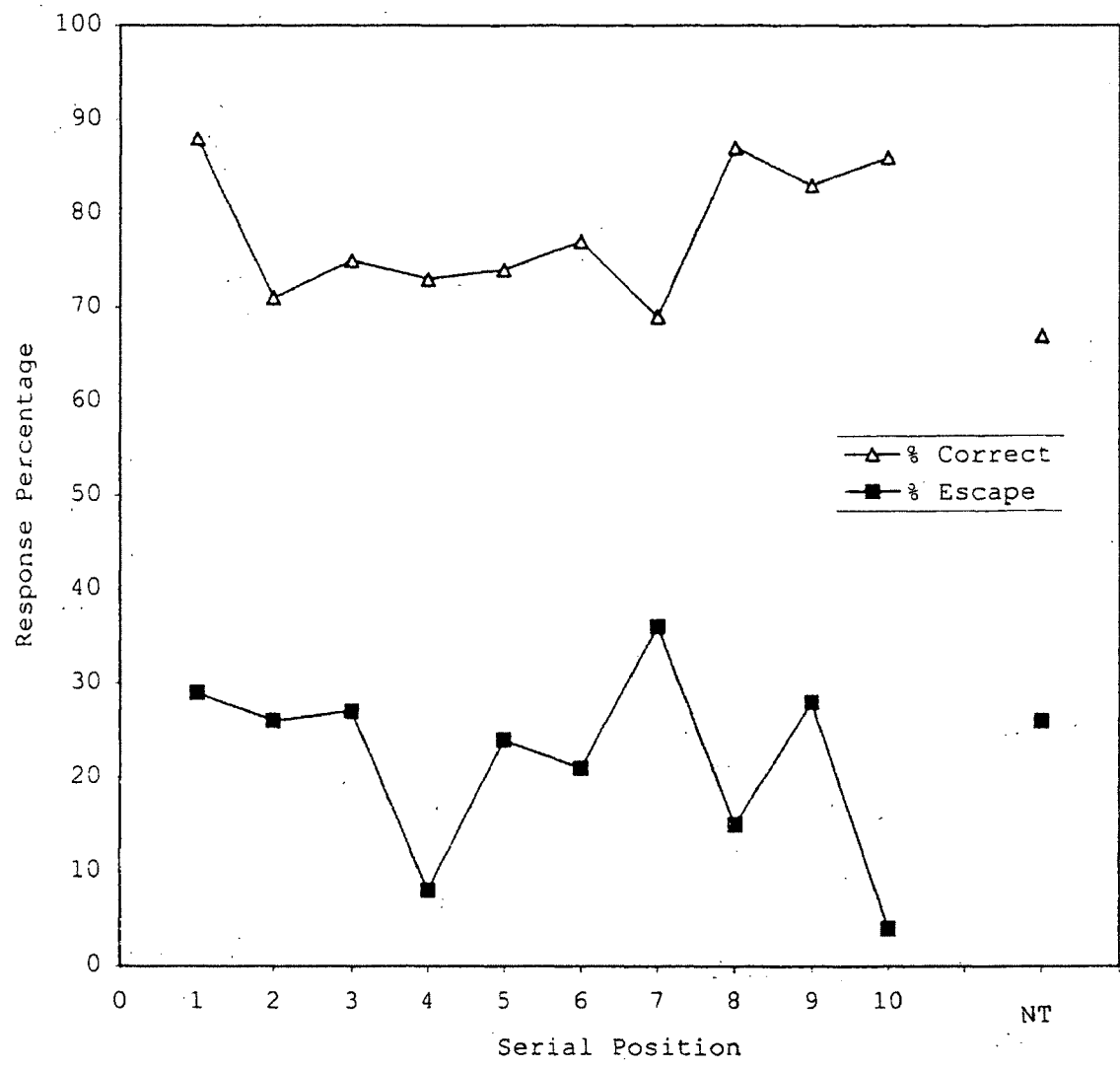


Figure 2B.

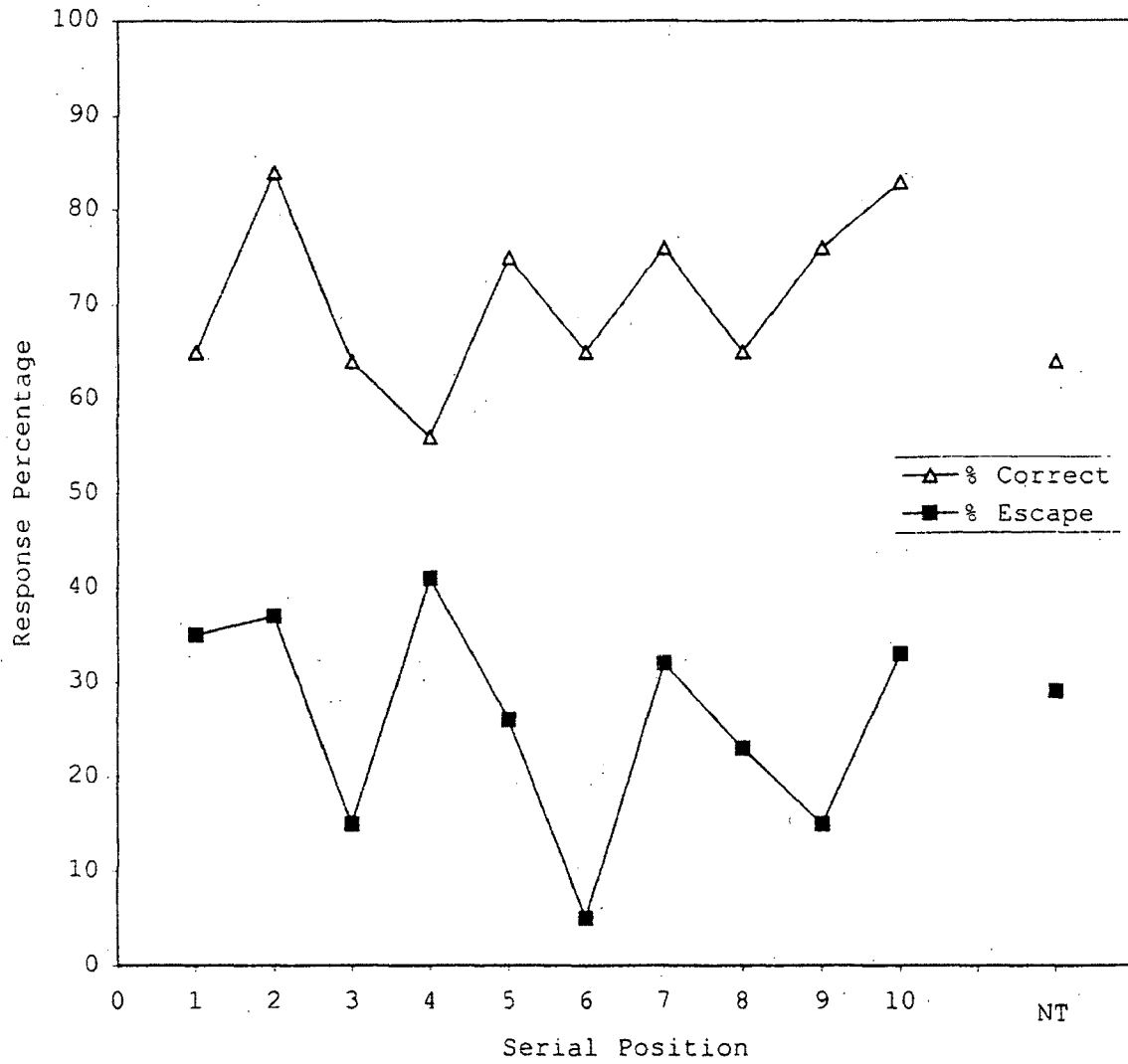


Figure 2C.

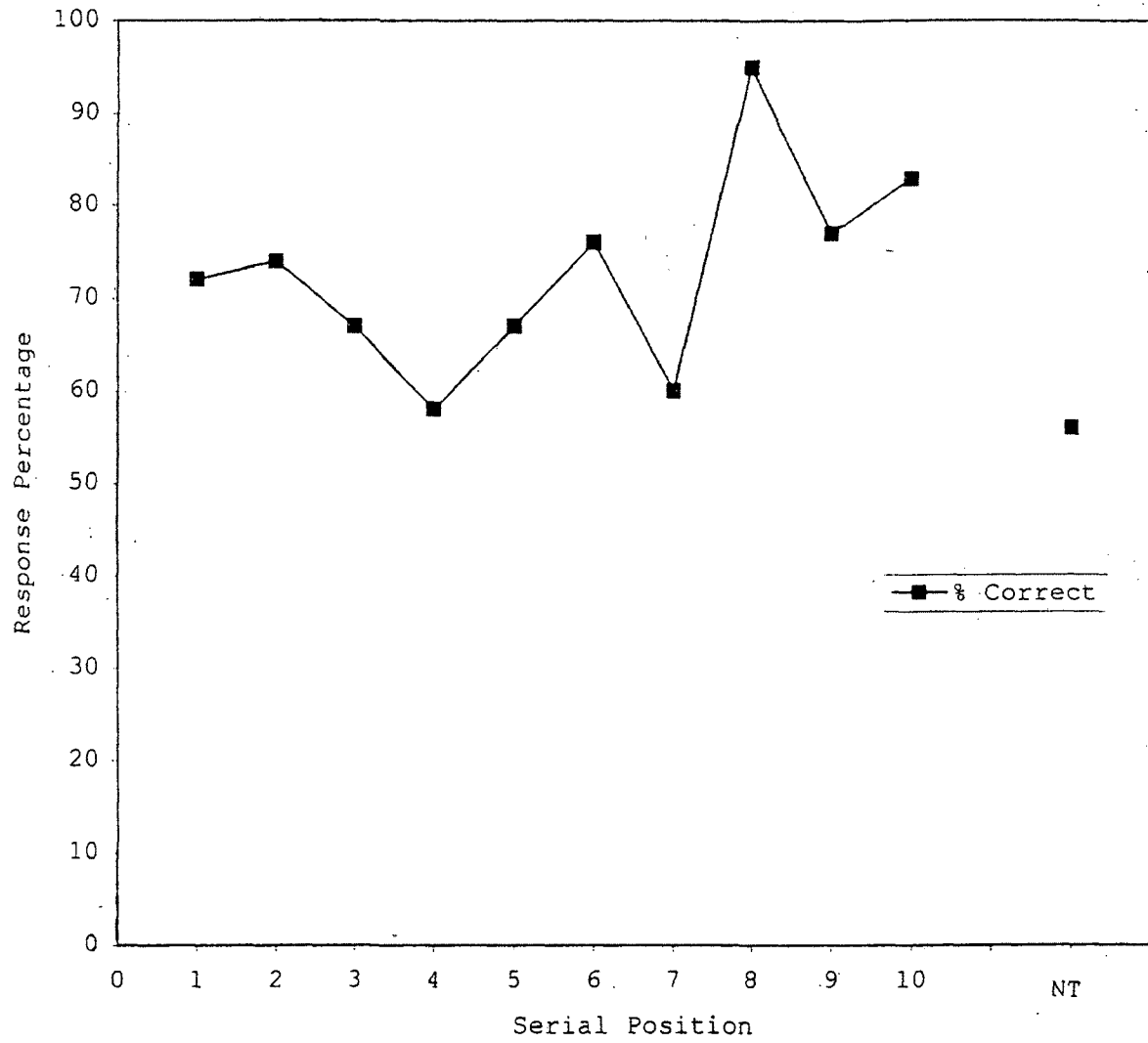


Figure 2D.

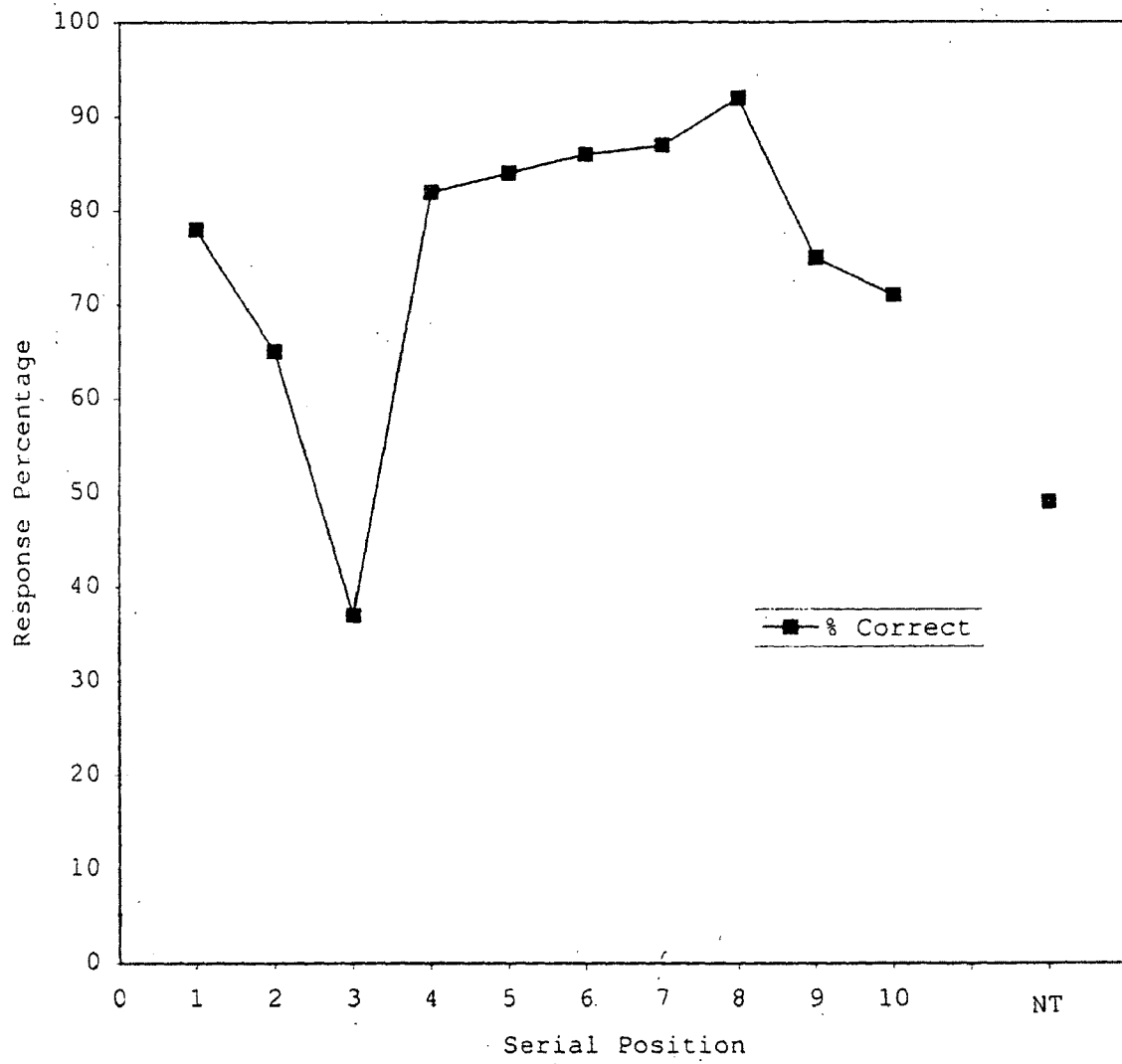


Figure 3.

